

Master Trace Sample

FINAL REPORT

This evaluation reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Joan M. Hill
Jason D. Machowski

Planning, Research, and
Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Intentionally Blank

CONTENTS

PURPOSE STATEMENT	iii
1. BACKGROUND	1
1.1 What is the Master Trace Sample (MTS) Database?	1
1.2 Motivations for Building the MTS Database	1
1.3 MTS Database Prototype Objectives	2
1.4 Intended Uses / Targeted Users	2
2. DESIGN AND METHODS	4
2.1 Database Description	4
2.2 Database Content	4
3. LIMITS	6
4. RECOMMENDATIONS	8
ACKNOWLEDGMENTS	9
REFERENCES	10

Intentionally Blank

PURPOSE STATEMENT

The Master Trace Sample database project merges Census 2000 data from multiple sources to provide information about cases in the various phases of data collection and processing. The objective of this effort is to support future methodological and operational analyses and decisions regarding the 2010 short form census. Over the last four years, the Planning, Research, and Evaluation Division staff has worked with subject-matter and database experts to create a complex, relational database for research purposes. This prototype database merges Census 2000 address frame, collection, enumeration, capture, processing, response, and coverage files. This merge yields a sophisticated database which allows quantitative insight into the relationship of key census processes (for example, how information travels from data collection through data processing). In addition to being an innovative research tool, the Census 2000 Master Trace Sample database is intended to serve as a model upon which we will improve in future censuses.

1. BACKGROUND

The Planning, Research, and Evaluation Division (PRED) has created a prototype Master Trace Sample database for 2010 short form census research. In 1999, PRED accepted the challenge of building a Census 2000 Master Trace Sample database with encouragement from stakeholders. Now, internal Census Bureau researchers are able to conduct explorations of operational and methodological issues aiming to benefit planners of the next decennial census that would not otherwise be possible.

1.1 What is the Master Trace Sample (MTS) database?

The MTS database contains a sample of Census 2000 housing unit records that will allow Census Bureau researchers to trace response and operational data, such as housing unit person count or field action codes, through stages of Census 2000 processing. These stages include address list development, data collection, data capture, and data processing. For the sample of housing unit records, the database contains all returns¹, which include housing unit level and person data.

The MTS database also contains data not typically analyzed in census evaluations. For example, the number of times an enumerator visited a housing unit during the Nonresponse Followup (NRFU) operation is contained in this database. In addition, this database links micro-level data such as enumerator production data with response data, which are not traditionally linked in census evaluations. An example of a research item that could make use of this database involves the correlation between repeated enumerator visits during NRFU and the rate of item non-response.

The purpose of the MTS database is to facilitate research on relationships among Census 2000 operations beyond the current Census 2000 Testing, Experimentation, and Evaluation Program.

1.2 Motivations for Building the MTS Database

“Instrumental in guiding methodological advances to be used in 2010 and beyond,” was how the National Research Council/National Academy of Sciences described an MTS database in *Measuring A Changing Nation* (1999). In this publication, they recommended that the Census Bureau create such a database. The Census Bureau had attempted to create a similar database in the previous decade based on a 1988 Council recommendation, but failed to do so due to operational difficulties and budget constraints.

In both recommendations, the National Academy of Sciences expressed interest in the creation of a database that would allow users to conduct a broad range of analyses on the quality of census

¹A return is a response to the Census, such as a mail return short form, an enumerator-filled form, or a response collected via the Internet or telephone. More than one return can be associated with a given housing unit ID.

content. The database would contain a sample of Census records selected in a way that users had the option of analyzing data at the block or interviewer level. In addition, users would be able to examine respondent data at various phases of data processing. This tool would be made available to users throughout the decade as fresh ideas for examining data quality and the census process arise.

Census managers with experience from previous censuses also encouraged the idea of building an MTS database. The possibilities of what census researchers could learn while planning the next census were appealing. The Decennial Management Division also supported the project with funding for headquarters staff and contractor support throughout the development phase of the project.

1.3 MTS Database Prototype Objectives

Census researchers will use the prototype MTS database to support future methodological and operational analyses and decisions related to the 2010 short form census.

While developing the database, staff met with upper management to ensure that the expectations of the Census Bureau and stakeholders (such as the National Academy of Sciences) were met. Staff also met with Census subject-matter experts to come up with research interests. Based on these broad objectives, the MTS team developed a series of functional requirements intended to cover a variety of research topics. However, we could not anticipate all research uses during the development phase of the project. For instance, Census Bureau researchers cannot trace every key measure through the census and similarly Census Bureau researchers cannot trace any one measure through all stages of data collection and processing. The team attempted to design the database to allow a variety of analyses, but it has specific limitations and cannot address every conceivable research question. See Section 3 for detailed limitations.

In addition, some of the ideas that the MTS team received during database development could not be incorporated into the prototype because of resource and timing limitations. However, these suggestions may greatly improve the usefulness of the MTS database and should be considered when designing future versions. See Section 4 for detailed recommendations.

1.4 Intended Uses / Targeted Users

With the wealth of data included in the MTS database, there is great potential for research in the following areas:

1. Modeling to identify and measure associations and relationships;
2. Tracing items, such as population count, through census processes; and
3. Investigating how to develop improved trace databases in future censuses.

For these research applications, the Census Bureau researcher can use the query capability or produce an extract for further analysis in statistical or database software.

1.4.1 Using the MTS for Multivariate Analyses versus Descriptive Statistics

The database is intended to be used to research hypotheses that involve relationships of various Census 2000 operations or systems. The MTS database is not intended to produce official totals or point estimates involving a single source file or program. The MTS estimates of descriptive statistics contain a level of uncertainty (that is, sampling error). However, descriptive results from single operations as reported for the full census do not contain sampling error. Consult the Census Bureau's website for information on Census operations as reported in the Census 2000 Testing, Experimentation, and Evaluation Program². For example, if you want descriptive statistics about Coverage Improvement Followup (CIFU), Census 2000 Testing, Experimentation, and Evaluation Program report I.4, *Coverage Improvement Followup* is a more appropriate source.

1.4.2 User Community is Specialized

The MTS database is intended to be used as an analysis tool for improving the programs and systems of the next Decennial Census. For this reason, as well as confidentiality issues, the MTS database access is limited to internal Census Bureau use. Census Bureau researchers interested in pursuing studies that will help guide the planning of the 2010 short form census will develop research proposals for review and approval by senior staff as well as planning groups guiding 2010 Census research.

²For more information on the Census 2000 Testing, Experimentation, and Evaluation Program, go to the following website: <http://www.census.gov/pred/www/>

2. DESIGN AND METHODS

Over the last four years, staff worked with internal subject-matter and database experts to design a complex relational database for research purposes. The MTS database design merges Census 2000 data from multiple sources, which allows quantitative insight into the relationship of key census processes (for example, how information travels from data collection through data processing).

2.1 Database Description

The MTS database links micro-level data from various stages of the Census 2000 process such as address frame development, data collection, data capture, data processing, and enumeration contact records. To facilitate research, data are linked at the following levels:

- Local Census Office (LCO),
- enumerator,
- housing unit,
- return (that is, census questionnaire),
- enumeration contact (that is, personal visit), and
- person.

The database contains a total of approximately 1.5 million Master Address File (MAF) housing unit identifiers (IDs) from a systematic ID sample and a block cluster sample. The block cluster sample contains all housing unit IDs within selected block clusters.

2.2 Database Content

The MTS database is intended to address a wide variety of research requests that link decennial census response, data collection, and processing information with enumeration characteristics. In general, the MTS database contains the following types of data:

- geography;
- census response data at various stages of processing;
- enumeration characteristics (related to operations and enumerators);
- record of contact information from the Nonresponse Followup (NRFU) and Coverage Improvement Followup (CIFU) forms;
- data capture system evaluation information from a reconciled keyed-from-image data set;
- geocoding error results from one of the Census 2000 evaluations; and
- housing unit status data (that is, occupied/vacant/delete/unresolved) from NRFU, CIFU and the Accuracy and Coverage Evaluation (A.C.E.).

The MTS database does not have Census 2000 person or housing unit coverage data from A.C.E. Coverage data represent cases that should have been added to or excluded from the Census based

on the A.C.E. Since the intent of the database is to trace existing census units through the process, coverage data are not included.

The database excludes special places and group quarters³. The database response data (that is, data provided by respondents in Census 2000) are limited to 100 percent item data and do not include sample data.⁴

³The MTS team initially considered including group quarters. However, to control the complexity of the task and reduce the number/types of sources to a more manageable level, the team decided to focus on the housing unit sample for the first prototype.

⁴ The 100 percent data are the data collected on all Census 2000 questionnaires, such as sex, age, race, and Hispanic origin. The sample data refer to items that appear only on the Census 2000 long form, such as educational attainment and income.

3. LIMITS

The following limitations apply to the overall MTS database:

- Any limitations present in the original Census 2000 files also are present in the MTS database, which contains data from numerous Census 2000 source files. Although the variables from these files went through testing to ensure that the data were properly extracted and merged, the values were not edited.⁵
- The MTS estimates of descriptive statistics contain a level of uncertainty (that is, sampling error). As previously stated in Section 1.4.1, the MTS database is intended to be used to research hypotheses that involve relationships of various Census 2000 operations or systems. The MTS database is not intended to produce official totals or point estimates involving a single source file or program. However, descriptive results from single operations as reported for the full census do not contain sampling error. Consult the Census Bureau's website for information on Census operations as reported in the Census 2000 Testing, Experimentation, and Evaluation Program.
- While the team attempted to design a database to handle a variety of issues, the database cannot address every conceivable research question. The research requirements are derived from a series of about 15 questions that represent anticipated research areas of high interest based on project goals and objectives. These functional requirements guided the design of the MTS database. In addition, there are ten supplementary questions that serve as a basis for building robustness and flexibility into the database. The MTS team attempted to create a database that could answer a wide variety of research questions about Census 2000. However, we could not anticipate all research uses during the development phase of the project.
- The database does not contain a comprehensive set of Census 2000 files. For example, the MTS database does not include the various Local Update of Census Addresses (LUCA) files or the bulk of the MAF extract files used to update the DMAF. The sources of data in the MTS database are intended to represent the major data collection, capture and processing steps for Census 2000.
- Enumerator characteristic and production data have limits. The association of enumerator data to a particular case is limited to the last enumerator who worked on the case. That enumerator is not necessarily associated with the full contact history of the case, if the case was worked by more than one enumerator. Another important limitation of these data is that they are "as reported" and "as keyed." Many of the limitations associated with enumerator files stem from the fact that the primary objective of these files was not

⁵ One exception is the record of contact data. These data were edited based on expert review of the enumerator records.

evaluation or research needs, but rather real time information for operational monitoring and tracking.

4. RECOMMENDATIONS

During the course of development, the MTS team received valuable input and suggestions from subject-matter and database experts. Some of these ideas could not be incorporated into the prototype because of complexity and resource/timing limitations. Developers who may refine the existing Census 2000 MTS prototype database or who may design future MTS databases should consider these suggestions. The primary recommendations are provided below.

Recommendations for Expanding the Census 2000 MTS Database

- The MTS could be expanded to include data on Group Quarters. Adding operational, response, and evaluation data associated with Group Quarters to the MTS database may prove valuable for planning the 2010 Census.
- Expanding the MTS database to include coverage measurement data associated with persons would provide an additional evaluation measure. The A.C.E. final Census Day housing unit status is the only Census 2000 A.C.E. data included in the prototype.

Recommendation for the 2010 Census MTS Database

- A formal evaluation should be implemented to assess both the usefulness of the database for research and the benefits to the Census Bureau of resulting products. This assessment should include quantitative and qualitative measures of effectiveness, accuracy, and value of MTS research products, as well as performance of the MTS database and Census Bureau researcher satisfaction. This information would guide the development of the 2010 MTS database.

Recommendation for Creating MTS Databases for other Censuses and Surveys

- The MTS concept of linking such valuable survey data into a comprehensive database for a specific survey or census, such as the American Community Survey, the Survey of Income and Program Participation, or the Economic Censuses, may prove useful. The Census 2000 MTS database links enumerator data to response data and quality measures for Census 2000. In the census, as in other major surveys, these key variables are traditionally located in separate systems or files and are not joined. Provided the proposed formal evaluation of the Census 2000 MTS finds the database useful, the Census Bureau may wish to consider building such a 'trace' database specific to each of its major surveys, as well as the Economic Censuses.

ACKNOWLEDGMENTS

The authors would like to thank and acknowledge the extraordinary contributions of the core Master Trace Sample Team: Mike Bentley, Julie Bouffard, Claude Jackson Jr., Lance Moore, Quang Phan, Cindy Rothhaas, Courtney Stapleton, George Train III., Frank Vitrano, and Henry Woltman.

Special thanks to the following groups and individuals:

Planning, Research, and Evaluation Division Management: Ruth Ann Killion, Debbie Bolton, James Treat, Randall Neugebauer

Demographic Statistical Methods Division: Dave Hubble (formerly a member of Planning, Research, and Evaluation Division Management)

Anteon Corporation: Ivan Abbott, Mark Kaplun, Carl Gorzynski Jr.

MicroStrategy: Nitin Tahiliani

Decennial Management Division

Other contributors: Florence Abramson, Leah Arnold, Cathy Ayoob, George Barnett, Joe Burcham, Danny Burkhead, Mike Carr, Joe Conklin, Judy Dawson, Ann Dimler, Debbie Fenstermaker, Debbie Griffin, Jennifer Guarino, Kevin Haley, Steve Hefter, Mark Holdrege, Dean Judson, Charles Kahn, Ellen Katzoff, Jay Keller, Fred Lestina Jr., Sue Love, George McLaughlin, Tracy Mattingly, Dan Philipp, Peggy Poole, Sherri Norris, Jennifer Reichert, Nancy Rogers, Meg Ruhnke, Jane Sandusky, Donna Souders, Dennis Stoudt, Laura Taylor, Cassandra Thomas

REFERENCES

National Research Council, *Measuring A Changing Nation: Modern Methods for the 2000 Census*. National Academy Press, Washington, D.C., 1999.

